INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Stability of a Polling Model with a Markovian Scheme

C. FRICKER
M.R. JAÏBI

N° 2278
Mai 1994

PROGRAMME 1

Rapport de recherche

Stabilité d'un modèle de polling à plan de travail Markovien

C. Fricker[a] and M.R. Jaïbi[b,*]

[a] INRIA, Domaine de Voluceau, BP.105, 78153 Le Chesnay Cedex, France

[b] Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

Abstract

Nous établissons la condition nécessaire et suffisante de stabilité d'un modèle de polling avec le plan de travail Markovien suivant. Le serveur sert les files selon une chaine de Markov bi-dimensionnelle, dont la première composante indique la file à servir et la seconde la politique de service à lui appliquer. Les politiques de services sont supposées "générales monotones", ce qui recouvre les politiques usuelles. Les arrivées aux files sont par groupes aux instants d'un processus de Poisson ; les durées de service aux files et les durées des déplacements entre files sont des suites indépendantes de variables indépendantes équidistribuées. Les démonstrations sont basées sur une propriéte de monotonie stochastique du système.

*This work was supported in part by a Fellowship of the Netherlands Organization for Scientific Research NWO-ECOZOEK.

Stability of a polling model with a Markovian scheme

C. Fricker[a] and M.R. Jaïbi[b,*]

[a] INRIA, Domaine de Voluceau, BP.105, 78153 Le Chesnay Cedex, France

[b] Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

Abstract

The necessary and sufficient condition for the stability of a polling model with the following Markovian scheme is established. The server attends to the queues according to a bi-dimensional Markov chain of which the first component indicates the queue to serve and the second the service policy to apply. The service policies are general monotonic and cover the usual ones. Batch arrivals to the queues occur at Poisson instants, the service times in the queues and the switchover times between the queues are independent i.i.d. sequences. The proofs are based on a stochastic monotonicity property of the system.

Keywords : polling system, Markovian scheme, multidimensional Markov chain, stochastic monotonicity, stability, local stability, heavy traffic, set-up times.

*This work was supported in part by a Fellowship of the Netherlands Organization for Scientific Research NWO-ECOZOEK.

1 Introduction and model description

A polling system, several queues attended to by a single server, is a model of distributed multiqueue systems sharing a single resource. There are many examples in communication, computer, production and road-traffic systems. A survey of the extensive literature on the subject is given in [9].

The main specifications of a polling model are the order in which the server attends to the queues, the service policies that are applied (a policy specifies the customers that are or may be served in a service period), and the statistical assumptions on the arrivals to the queues, the required service times and the switchover times of the server between the queues. One of the fundamental issues is then the derivation of the stability conditions. In two recent papers, this is done by different approaches for the following specifications: [4] for the periodic routing and general monotonic service policies, [2] for the Markovian routing and a fixed limited gated service policy at each queue, both with arrivals at Poisson times and independent sequences of i.i.d. service times and switchover times. (See also [1, 3, 5, 6, 7]). In this paper we apply the approach based on monotonicity arguments used in [4] to derive the necessary and sufficient condition for the stability of the model where the server attends to the queues according to a bi-dimensional Markov chain of which the first component indicates the queue to serve and the second the service policy to apply. The service policies are general monotonic, as specified below, and cover the usual ones like for example the exhaustive and the gated policies in their pure and in their limited versions. This scheme generalizes those in the two papers mentioned above. The randomization of the selection of the service policies and the fact that this selection affects the subsequent scheme may be interesting features for optimization purposes.

The polling system is composed of c queues, labelled 1 through c, successively visited by the server. A sequence of δ general monotonic service policies, referred to by f_1 through f_δ and not necessarily mutually different, specifies the possible service policies. The working scheme of the server is determined by a Markov chain $(X_n, Y_n)_n$ with values in $\{1, \ldots, c\} \times \{1, \ldots, \delta\}$, transition matrix $(r_{(j,\alpha),(k,\beta)})$ and unique invariant distribution $(p_{(j,\alpha)})$ ($p_{(j,\alpha)} \geq 0$, $\sum_\alpha p_{(j,\alpha)}) > 0$, $\sum_{(j,\alpha)} p_{(j,\alpha)}) = 1$), where $(X_n, Y_n) = (j, \alpha)$ means that visit n is to queue j according to policy f_α. At completion of visit n, a transition from (j, α) to (k, β) induces a switchover time $s_{(j,\alpha),(k,\beta),n}$ of the server. The sequence $(s_{(j,\alpha),(k,\beta),n})_n$ is i.i.d. with finite mean $S_{(j,\alpha),(k,\beta)} \geq 0$ and a general distribution. The expected stationary switchover time per visit, when (X_n, Y_n) is stationary, is thus $\bar{S} = \sum_{(j,\alpha),(k,\beta)} p_{(j,\alpha)} r_{(j,\alpha),(k,\beta)} S_{(j,\alpha),(k,\beta)}$ and is assumed to be positive. Customers arrive to the queues in groups of i.i.d. joint sizes at the instants of a Poisson process (when the group sizes add up to unity, the arrival processes are independent Poisson processes). The arrival process N_j to queue j is thus compound Poisson with arrival rate $\lambda_j = E(N_j(0, 1]) > 0$, where $N_j(u, v]$ is the number of arrivals in the time interval $(u, v]$. The service times $(\sigma_j^m)_m$ required in queue j are i.i.d. with a general distribution and finite mean σ_j. For each j we assume $\rho_j = \lambda_j \sigma_j < 1$ to ensure the stability of queue j when it operates as a standard M/G/1 queue in isolation. The Markov chain $(X_n, Y_n)_n$, the sequences of switchover times and of service times and the arrival process are assumed to be mutually independent.

Under the above assumptions, we prove that the polling system is stable, i.e. it admits a stationary regime with integrable regeneration times, if and only if

$$\mathcal{C}: \qquad \hat{\rho} + \max_{1\le j\le c} \left(\frac{\lambda_j}{\sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *})} \right) \bar{S} < 1$$

where $\hat{\rho} = \sum_{j=1}^{c} \rho_j$ is the total traffic load of the system, $E(F^{\alpha *})$ is the maximum expected number of served customers in a service period according to policy f_α and is defined below ($0 < E(F^{\alpha *}) \le \infty$ and, by convention, $0 \times \infty = 0$ and $\lambda_j/\infty = 0$) and thus $\sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *})$ is the maximum expected number of served customers in queue j per visit when (X_n, Y_n) is stationary.

A general monotonic service policy and the associated maximum expected number of served customers are defined as follows. Let a service policy be applied to a queue with a compound Poisson arrival process N with rate λ, and i.i.d. service times $(\sigma^l)_l$ with mean σ. We describe a service period started with x customers by
the (random) number $f(x)$ of customers served in the period,
the duration $v(x)$ of the period, and
the queue length $\varphi(x)$ at the end of the period.
The policy is general monotonic if the following three properties hold. First, there is work-conservation, so that

$$\begin{aligned} v(x) &= \sum_{l=1}^{f(x)} \sigma^l, \\ \varphi(x) &= x - f(x) + N(0, v(x)], \end{aligned}$$

with $f(0) = v(0) = \varphi(0) = 0$. Second, the selection of a customer for service is done independently of its particular service time and of the past up to the start of the service period. Thus the distribution of (f, v, φ) does not depend on the order in which the customers are served, and Wald's formula applies to calculate the expectations:

$$\begin{aligned} E(v(x)) &= \sigma E(f(x)), \\ E(N(0, v(x)]) &= \lambda \sigma E(f(x)). \end{aligned}$$

By these two properties, the distribution of (f, v, φ) is uniquely generated by and characterizes the policy. Third, the policy is monotonic in the sense that as the queue size grows, the number of customers served in a service period grows stochastically, but such that the number of customers left at the end of the period grows stochastically as well. In other terms,

$$(f(x), \varphi(x)) \textit{ is } \le_d\textit{-monotone in } x$$

where $\le_d$-monotone refers to the stochastic monotonicity or monotonicity in distribution (cf. [8]). It follows that $(f(x), v(x), \varphi(x))$ is $\le_d$-monotone in x and that, as x goes to ∞, $(f(x), v(x))$ converges in distribution to a (possibly degenerate) random vector (F^*, V^*), and obviously $(f(x), v(x)) \le_d (F^*, V^*)$ for any x. In fact, F^* and V^* would be the number of customers served in a service period and the duration of the service period if there were infinitely many customers waiting at the start of the service period, respectively, with

$E(V^*) = \sigma E(F^*) \leq \infty$. The integrability of these stochastic bounds plays an important role in the analysis and leads to the following classification of the service policy to which we refer by policy f:

Policy f is said to be of limited type if its bounds F^* and V^* are integrable, and of unlimited type otherwise.

The three properties are satisfied by the usual policies. The exhaustive policy, for which $f(x)$ is the number of customers served until the queue is emptied, and the gated policy, for which $f(x) = x$, are of unlimited type, both with $F^* \equiv \infty$. The limited versions of these policies, where the server cannot serve more than a (random) number L of customers in a service period, admit the bound $F^* = L$. When L has a geometric distribution, we have the Bernoulli versions which are of limited type. Other examples are the decrementing policies and the time limited policy without preemption which are easily classified (cf. [4]).

The paper is organized as follows. In section 2, the monotonicity property of the system is exhibited. The stability condition is established and some extensions of the result are stated in section 3.

2 Monotonicity

Let T_n ($n \geq 1$) be the n-th polling time, the time at which the n-th visit starts ($T_1 = 0$). Let $Q_j(T_n)$ be *the number of customers in queue j at time T_n* and

$$M_n = (Q_1(T_n), \ldots, Q_c(T_n)).$$

We use the shorthand notation $Q_n = Q_{X_n}(T_n)$ for the queue X_n served in visit n and s_n for the n-th switchover time $s_{(X_n,Y_n),(X_{n+1},Y_{n+1}),n}$. The characteristics of visit n are
the number F_n of served customers,
the duration V_n of the visit, and
the number Φ_n of customers left in queue X_n at the end of the visit.
These quantities satisfy

$$\begin{aligned} T_{n+1} - T_n &= V_n + s_n\,, & (1)\\ Q_j(T_{n+1}) - Q_j(T_n) &= N_j(T_n, T_n + V_n + s_n] - F_n\,\delta_{j,X_n} & (2) \end{aligned}$$

where $\delta_{j,k}$ is the Kronecker symbol. By the Poisson and independence assumptions and by the properties imposed on the service policies, the sequence $(X_n, Y_n, M_n)_n$ is a Markov chain which is irreducible on $\{(j,\alpha),\ p_{(j,\alpha)} > 0\} \times \mathbb{N}^c$. The conditional distribution of (F_n, V_n, Φ_n) given $(X_n, Y_n) = (j, \alpha)$ and $Q_n = x$ is the distribution of $(f_\alpha(x), v_\alpha(x), \varphi_\alpha(x))$ generated by policy f_α, i.e.

$$(F_n, V_n, \Phi_n)\ |_{X_n=j,Y_n=\alpha,Q_n=x} \stackrel{d}{=} (f_\alpha(x), v_\alpha(x), \varphi_\alpha(x)).$$

Moreover

$$\begin{aligned} E(1_{\{X_n=j\}}V_n) &= \sigma_j E(1_{\{X_n=j\}}F_n), & (3)\\ E(1_{\{X_n=j\}}N_k(T_n, T_n + V_n]) &= \lambda_k \sigma_j E(1_{\{X_n=j\}}F_n). & (4) \end{aligned}$$

Let $(\nu_{(j,\alpha),n})_n$ be the successive passages of $(X_\cdot, Y_\cdot)$ to (j,α), or visits to queue j according to policy f_α. A (j,α)-cycle is the period of time between two consecutive visits to (j,α). We assume without loss of generality that $p_{(1,1)} > 0$ and we mainly consider $(1,1)$-cycles. In that case, we omit the reference to $(1,1)$ and write ν_n in place of $\nu_{(1,1),n}$. When necessary, the calculations and results are adapted in a straightforward way to (j,α)-cycles. Thus, cycle n takes place between instants T_{ν_n} and $T_{\nu_{n+1}}$. The number $G_{j,n}$ of customers served in queue j in cycle n is

$$G_{j,n} = \sum_{l=\nu_n}^{\nu_{n+1}-1} 1_{\{X_l=j\}} F_l,$$

and

$$\begin{aligned} T_{\nu_{n+1}} - T_{\nu_n} &= \sum_{l=\nu_n}^{\nu_{n+1}-1} (V_l + s_l), \\ Q_j(T_{\nu_{n+1}}) - Q_j(T_{\nu_n}) &= N_j(T_{\nu_n}, T_{\nu_{n+1}}] - G_{j,n}. \end{aligned} \tag{5}$$

The restricted Markov chain $(M_{\nu_n})_n$ is homogeneous, aperiodic and irreducible on $\mathbb{N}^c$ and satisfies the following crucial monotonicity property.

Proposition 1 *Suppose $(X_1, Y_1) = (1,1)$ and $M_1 = (0,\ldots,0)$. Then $(M_{\nu_n})_n$ is $\leq_d$-monotone. For each j, $E(G_{j,n})$ is non-decreasing in n and is bounded by*

$$\bar{G}_j^* = \frac{1}{p_{(1,1)}} \sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *}).$$

Remarks. 1. $\bar{G}_j^*$ is the maximum expected number of customers that can be served in queue j in a cycle. Note that $\bar{G}_j^* = \infty$ if, for one α such that $p_{(j,\alpha)} > 0$, policy f_α is of unlimited type ($E(F^{\alpha *}) = \infty$). Otherwise $\bar{G}_j^* < \infty$ (by convention, $0 \times \infty = 0$).
2. For (j,α)-cycles, Proposition 1 states that if $(X_1, Y_1) = (j,\alpha)$ and $M_1 = (0,\ldots,0)$, then $(M_{\nu_{(j,\alpha),n}})_n$ is $\leq_d$-monotone and for each k, the expected number $E(G_{k,(j,\alpha),n})$ of customers served in queue k in (j,α)-cycle n is non-decreasing in n and is bounded by

$$\bar{G}_{k,(j,\alpha)}^* = \frac{1}{p_{(j,\alpha)}} \sum_{\beta=1}^{\delta} p_{(k,\beta)} E(F^{\beta *}).$$

Proof. Let $\pi_{(j,\alpha),(k,\beta)}$ be the operator defined by

$$\pi_{(j,\alpha),(k,\beta)} h(m) = E\left(h(M_{n+1}) \mid X_n = j, Y_n = \alpha, X_{n+1} = k, Y_{n+1} = \beta, M_n = m\right)$$

where $m = (m_1, \ldots, m_c) \in \mathbb{N}^c$ and h is any real function defined on $\mathbb{N}^c$ for which the expectation exists. We first prove that $\pi_{(j,\alpha),(k,\beta)}$ is $\leq_d$-monotone, which holds if $\pi_{(j,\alpha),(k,\beta)} h$ is $\leq$-monotone when h is $\leq$-monotone, where $\leq$ refers to the componentwise partial order on $\mathbb{N}^c$ (cf [8]). For notational ease, consider $\pi_{(1,\alpha),(k,\beta)}$. From equations (1)-(2) and by the Poisson and independence assumptions, we have

$$\begin{aligned} \pi_{(1,\alpha),(k,\beta)} h(m) = & \int_s d\, P_{s_{(1,\alpha),(k,\beta)}}(s) \\ & E\left(h\left(\varphi_\alpha(m_1) + N_1(v_\alpha(m_1), v_\alpha(m_1) + s], m_2 + N_2(0, v_\alpha(m_1) + s], \ldots\right)\right) \end{aligned}$$

because the conditional distributions of (V_n, Φ_n) and of s_n given $\{X_n = 1, Y_n = \alpha, X_{n+1} = k, Y_{n+1} = \beta, M_n = m\}$ are those of $(v_\alpha(m_1), \varphi_\alpha(m_1))$ and $s_{(1,\alpha),(k,\beta)}$, respectively. The argument of the function h is $\leq_d$-monotone in m because it is the sum of two independent random vectors which are $\leq_d$-monotone: $(\varphi_\alpha(m_1), m_2 + N_2(0, v_\alpha(m_1))], \ldots)$ because $(f_\alpha, v_\alpha, \varphi_\alpha)$ is $\leq_d$-monotone and $(N_1(v_\alpha(m_1), v_\alpha(m_1) + s], N_2(v_\alpha(m_1), v_\alpha(m_1) + s], \ldots)$ because its distribution does not depend on m. Thus $\pi_{(1,\alpha),(k,\beta)}h$ is $\leq$-monotone if h is $\leq$-monotone, where $\leq$ refers to the componentwise partial order on $I\!N^c$. It implies that $\pi_{(1,\alpha),(k,\beta)}$ is $\leq_d$-monotone, and so is $\pi_{(j,\alpha),(k,\beta)}$ for each j. Consider now the transition operator $\tilde{\pi}$ of the Markov chain $(M_{\nu_n})_n$. It is the weighted average, calculated over all possible trajectories of $(X_., Y_.)$ in a cycle, of products of elementary operators $\pi_{(j,\alpha),(k,\beta)}$, and is thus $\leq_d$-monotone. Consequently, if $M_{\nu_n} \leq_d M_{\nu_{n+1}}$, then $M_{\nu_{n+1}} \leq_d M_{\nu_{n+2}}$. On the other hand, if $(X_1, Y_1) = (1,1)$ and $M_1 = M_{\nu_1} = (0, \ldots, 0)$, obviously $M_{\nu_1} \leq_d M_{\nu_2}$. By immediate induction, $(M_{\nu_n})_n$ is $\leq_d$-monotone.

The second assertion follows by a coupling argument. Consider two versions of the polling system such that $M_1 \leq_d \tilde{M}_1$. Consider the first cycle of each version coupled by the same trajectories of $(X_., Y_.)$: the queues are served in the same order according to the same policies. Let $A_{m,(x_.,y_.)} = \{\nu_2 = m, (X_l, Y_l) = (x_l, y_l), 1 \leq l \leq m\}$ be a possible trajectory. Given $A_{m,(x_.,y_.)}$, for each $l \leq m$ we have $M_l \mid_{A_{m,(x_.,y_.)}} \leq_d \tilde{M}_l \mid_{A_{m,(x_.,y_.)}}$ because the operators $\pi_{(j,\alpha),(k,\beta)}$ are $\leq_d$-monotone. Because the service policies are monotonic, the inequality holds in distribution and thus in expectation for the respective numbers of served customers: $E(F_l \mid A_{m,(x_.,y_.)}) \leq E(\tilde{F}_l \mid A_{m,(x_.,y_.)})$. On the other hand,

$$E(G_{j,1}) = \sum_{m,(x_.,y_.)} P(A_{m,(x_.,y_.)}) \sum_{\alpha=1}^{\delta} \sum_{l=1}^{m-1} E(1_{\{(X_l,Y_l)=(j,\alpha)\}} F_l | A_{m,(x_.,y_.)}),$$

and it easily follows that $E(G_{j,1}) \leq E(\tilde{G}_{j,1})$. Moreover, for $l < m$

$$E(1_{\{(X_l,Y_l)=(j,\alpha)\}} F_l | A_{m,(x_.,y_.)}) \leq E(1_{\{(X_l,Y_l)=(j,\alpha)\}} | A_{m,(x_.,y_.)}) E(F^{\alpha *}).$$

Thus

$$\begin{aligned}
E(G_{j,1}) &\leq \sum_{m,(x_.,y_.)} P(A_{m,(x_.,y_.)}) \sum_{\alpha=1}^{\delta} \sum_{l=1}^{m-1} E(1_{\{(X_l,Y_l)=(j,\alpha)\}} \mid A_{m,(x_.,y_.)}) E(F^{\alpha *}) \\
&\leq \sum_{\alpha=1}^{\delta} E(\sum_{l=1}^{\nu_2 - 1} 1_{\{X_l = j, Y_l = \alpha\}}) E(F^{\alpha *}) \\
&= \sum_{\alpha=1}^{\delta} \frac{p_{(j,\alpha)}}{p_{(1,1)}} E(F^{\alpha *}) \\
&= \bar{G}_j^*
\end{aligned}$$

whatever the distribution of M_{ν_1} is. Now, in the system, the trajectories of $(X_., Y_.)$ within different cycles are independent. Thus, by the coupling arguments above, $M_{\nu_n} \leq_d M_{\nu_{n+1}}$ implies that $E(G_{j,n}) \leq E(G_{j,n+1})$ and $E(G_{j,n})$ is bounded by $\bar{G}_j^*$. □

We now define subsystems of the polling system obtained by the saturation of some queues. Without loss of generality, we suppose from here on that the c queues are numbered such that:

- *the queues* $1,\dots,b$ *are eventually served according to a policy of unlimited type* ,
- *the queues* $b+1,\dots,c$ *are always served according to policies of limited type and are numbered such that* $\lambda_j/\sum_{\alpha=1}^{\delta} p_{(j,\alpha)}E(F^{\alpha*})$ *is non-decreasing in* j.

Thus we have $j \leq b$ iff there exists α such that $p_{(j,\alpha)} > 0$ and policy f_α is of unlimited type, and in that case $\lambda_j/\sum_{\alpha=1}^{\delta} p_{(j,\alpha)}E(F^{\alpha*}) = 0$. The cases $b = 0$ or $b = c$, where all policies are of the same type, are not excluded.

A queue is said to be saturated if there are infinitely many customers waiting in it. If queue j with $j \leq b$ is saturated at time 0, the duration of the first visit according to a service policy of unlimited type is non-integrable if not infinite: this will be seen to exclude stability. If $j > b$, only policies of limited type are involved in the service of queue j. At each visit to the queue, the number of served customers and the duration of the visit have the distributions of the bounds of the applied service policy and are integrable. Moreover, for each policy the durations of the visits are i.i.d. random variables, independent of all quantities relative to the other queues and of all switchover times, just like the latter ones. Thus, queue j stays saturated for ever but the polling system works. The others queues constitute now a polling (sub)system in which additional "switchover" times result from the saturation of queue j. Several queues may be saturated as long as no policy of unlimited type is involved.

Let $\mathcal{S}$ refer to the polling system. Let $\mathcal{S}^e$ ($b \leq e \leq c$) be the subsystem formed by the queues $1,\dots,e$ resulting from the saturation of the queues $e+1,\dots,c$, with the same Markovian scheme (X_n, Y_n). More precisely, when $X_n = j \leq e$ and $Y_n = \alpha$, queue j is served according to policy f_α as in $\mathcal{S}$. When $X_n = j > e$ and $Y_n = \alpha$, no queue of $\mathcal{S}^e$ is served but the server is subject to two consecutive switchover times, $V_n^{\alpha*}$ having the distribution of the bound $V^{\alpha*}$ of policy f_α followed by the switchover time s_n. However, in order to facilitate the comparison of the subsystems, we keep the periods of unavailability $V_n^{\alpha*}$ separate from the original switchover times and we still consider their beginnings as imaginary "polling instants" of the polling system $\mathcal{S}^e$.

The state of subsystem $\mathcal{S}^e$ at the n-th polling instant T_n^e is given by the Markov chain (X_n, Y_n, M_n^e), where $M_n^e = (Q_1^e(T_n^e),\dots,Q_e^e(T_n^e))$ and $Q_j^e(t)$ is the number of customers in queue j in $\mathcal{S}^e$ at time t. We have

$$T^e_{\nu_{n+1}} - T^e_{\nu_n} = \sum_{l=\nu_n}^{\nu_{n+1}-1} \left(1_{\{X_l \leq e\}} V_l + 1_{\{X_l > e\}} V_l^{Y_l *} + s_l\right), \tag{6}$$

$$Q_j^e(T^e_{\nu_{n+1}}) - Q_j^e(T^e_{\nu_n}) = N_j(T^e_{\nu_n}, T^e_{\nu_{n+1}}] - G^e_{j,n} \qquad (j \leq e), \tag{7}$$

where $G^e_{j,n}$ is the number of customers served in queue j in cycle n in $\mathcal{S}^e$. Our previous calculations and results are easily adapted to $\mathcal{S}^e$. In particular, if $(X_1, Y_1) = (1,1)$ and $M_1^e = (0,\dots,0)$, $M^e_{\nu_n}$ is $\leq_d$-monotone and $E(G^e_{j,n})$ is non-decreasing and bounded by $\bar{G}_j^*$. Moreover we have the following dominance property, where $M^{g|e}$ denotes the e first components of a vector M^g having $g > e$ components.

Lemma 1 *For* $e < g$ *both in* $\{b,\dots,c\}$, $\mathcal{S}^e$ *dominates* $\mathcal{S}^g$ *in the sense that if* $M_1^{g|e} \leq_d M_1^e$, *then* $M_n^{g|e} \leq_d M_n^e$ *for all* n.

The proof is by coupling the two subsystems by the same trajectories of $(X_., Y_.)$, as done in the proof of Proposition 1, and it is provided for the case of deterministic periodic (X_n, Y_n) in [4]. It is therefore omitted.

3 Stability

The polling system is said to be *stable* if it admits a stationary regime with integrable regeneration times. Our criterion is the existence of a proper stationary distribution for the Markov chain (M_{ν_n}) such that the mean cycle time is finite.

Let us recall the condition stated in the introduction,

$$\mathcal{C}: \qquad \hat{\rho} + \max_{1\leq j\leq c}\left(\frac{\lambda_j}{\sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *})}\right)\bar{S} < 1,$$

where $\bar{S} = \sum_{(j,\alpha),(k,\beta)} p_{(j,\alpha)} r_{(j,\alpha),(k,\beta)} S_{(j,\alpha),(k,\beta)}$ is the expected stationary switchover time in $\mathcal{S}$. For subsystem $\mathcal{S}^e$, the total traffic load is $\hat{\rho}^e = \sum_{j=1}^{e} \rho_j$, the expected stationary switchover time is

$$\bar{S}^e \;=\; \bar{S} + \sum_{k=e+1}^{c} \sigma_k \sum_{\alpha=1}^{\delta} p_{(k,\alpha)} E(F^{\alpha *})$$

and the condition corresponding to $\mathcal{C}$ is

$$\mathcal{C}^e: \qquad \hat{\rho}^e + \frac{\lambda_e}{\sum_{\alpha=1}^{\delta} p_{(e,\alpha)} E(F^{\alpha *})}\bar{S}^e < 1.$$

Indeed, the queues have been numbered such that

$$\frac{\lambda_e}{\sum_{\alpha=1}^{\delta} p_{(e,\alpha)} E(F^{\alpha *})} = \max_{1\leq j\leq e} \frac{\lambda_j}{\sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *})}.$$

Moreover it is easy to see that $\mathcal{C}^{e+1}$ implies $\mathcal{C}^e$ $(b \leq e < c)$. Because $\sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *}) = \infty$ for $j \leq b$, $\mathcal{C}^b$ reduces to $\hat{\rho}^b < 1$ and $\mathcal{C}^c$ coincides with $\mathcal{C}$.

By assumption, $\rho_j < 1$ for $1 \leq j \leq c$. It ensures the integrability of T_n^e, $Q_j^e(T_n^e)$, F_n^e and V_n^e for all n and j. If we take the expectations in equations (6)- (7) and if we use (3)-(4), we get

$$E\left(T^e_{\nu_{n+1}} - T^e_{\nu_n}\right) \;=\; \sum_{j=1}^{e} \sigma_j E(G^e_{j,n}) + S^e, \tag{8}$$

$$E\left(Q^e_j(T^e_{\nu_{n+1}}) - Q^e_j(T^e_{\nu_n})\right) \;=\; \lambda_j\left(\sum_{j=1}^{e} \sigma_j E(G^e_{j,n}) + S^e\right) - E(G^e_{j,n}), \tag{9}$$

where $S^e = p^{-1}_{(1,1)}\bar{S}^e$ is the mean of the total switchover time in a cycle in $\mathcal{S}^e$. Under the condition $\mathcal{C}^e$, the mean numbers of served customers in $\mathcal{S}^e$ do not reach their maxima and the queue lengths do not converge in distribution to infinity, as shown by the following lemmas.

Lemma 2 *Suppose* $(X_1, Y_1) = (1, 1)$ *and* $M_1^e = (0, \ldots, 0)$. *If condition* C^e *holds, then* $\bar{G}_j^e = \lim_{n\to+\infty} E(G_{j,n}^e) < \bar{G}_j^*$ *for each* $j \leq e$, *with* $\bar{G}_j^*$ *infinite for* $j \leq b$ *and finite for* $j > b$.

Proof. We provide the proof for $e = c$ and omit the superscript. By Proposition 1, M_{ν_n} is $\leq_d$-monotone. Thus for each j and each n

$$E\left(Q_j(T_{\nu_{n+1}}) - Q_j(T_{\nu_n})\right) \geq 0$$

which inserted in (9) provides

$$E(G_{j,n}) \leq \lambda_j \left(\sum_{k=1}^{c} \sigma_k E(G_{k,n}) + S\right) \qquad (1 \leq j \leq c). \tag{10}$$

If we multiply both sides by σ_j and sum up on j from 1 to b, we get

$$(1 - \hat{\rho}^b) \sum_{j=1}^{b} \sigma_j E(G_{j,n}) \leq \hat{\rho}^b \left(\sum_{j=b+1}^{c} \sigma_j E(G_{j,n}) + S\right). \tag{11}$$

By Proposition 1, $\bar{G}_j = \lim_{n\to\infty} E(G_{j,n})$ exists. If we take limits in (11), it is easy to see that if $\sum_{j=1}^{b} \sigma_j \bar{G}_j = \infty$, then $\hat{\rho}^b \geq 1$ because the right hand side is bounded by $\hat{\rho}^b \left(\sum_{j=b+1}^{c} \sigma_j \bar{G}_j^* + S\right) < \infty$. Thus $\hat{\rho}^b < 1$ implies that $\bar{G}_j < \infty = \bar{G}_j^*$ for each $j \leq b$. Moreover, inequalities (10) and (11) hold for the limits, and their combination provides

$$(1 - \hat{\rho}^b)\bar{G}_j \leq \lambda_j \left(\sum_{k=b+1}^{c} \sigma_k \bar{G}_k + S\right) \qquad (b < j \leq c).$$

When $\hat{\rho} < 1$, which is implied by C, straightforward but tedious algebra leads to

$$(1 - \hat{\rho}^j)\bar{G}_j \leq \lambda_j \left(\sum_{k=j+1}^{c} \sigma_k \bar{G}_k + S\right) \qquad (b < j \leq c). \tag{12}$$

Because $\bar{G}_k \leq \bar{G}_k^* = p_{(1,1)}^{-1} \sum_{\alpha=1}^{\delta} p_{(k,\alpha)} E(F^{\alpha *})$ by Proposition 1, for each $b < j \leq c$, the corresponding inequality of (12) implies that

$$\begin{aligned}
(1 - \hat{\rho}^j)\bar{G}_j &\leq p_{(1,1)}^{-1} \lambda_j \left(\sum_{k=j+1}^{c} \sigma_k \sum_{\alpha=1}^{\delta} p_{(k,\alpha)} E(F^{\alpha *}) + \bar{S}\right) \\
&= p_{(1,1)}^{-1} \lambda_j \bar{S}^j.
\end{aligned}$$

On the other hand, C, which holds by assumption, implies C^j which is equivalent to $\lambda_j \bar{S}^j < (1 - \hat{\rho}^j)\, p_{(1,1)}\, \bar{G}_j^*$. Thus $\bar{G}_j < \bar{G}_j^*$ for each $j > b$. □

Lemma 3 *Suppose that condition* C^e *holds, and that* $(X_1, Y_1) = (1, 1)$ *and* $M_1^e = (0, \ldots, 0)$. *Then,* $(Q_j^e(T_{\nu_n}^e))_n$ *has a limit in distribution which is proper for* $j \leq b$ *and is not concentrated at infinity for* $b < j \leq e$.

Proof. Again we provide the proof for $e = c$ and omit the superscript.

Define $\tau^l_{j,n}$ and $\tau^l_{(j,\alpha),n}$ as the l-th passages after ν_n of X_n to j (irrespective of the value of Y_n) and of (X_n, Y_n) to (j,α), respectively ($\tau^1_{1,n} = \nu_n$). Consider a stationary event relative to the trajectory of $(X_\cdot, Y_\cdot)$ in a cycle, and let A_n be the event in cycle n ($P(A_n) = P(A_1) > 0$). Because all visits to queue j after ν_n are counted by $\tau^l_{j,n}$, we have

$$Q_j(T_{\tau^1_{j,n}})1_{A_n} \geq Q_j(T_{\nu_n})1_{A_n} \tag{13}$$

and for each $l > 0$

$$Q_j(T_{\tau^{l+1}_{j,n}})1_{A_n} \quad \geq \quad \left(Q_j(T_{\tau^l_{j,n}}) - F_{\tau^l_{j,n}}\right)1_{A_n}. \tag{14}$$

By Proposition 1, $(Q_j(T_{\nu_n}))_n$ is $\leq_d$-monotone and therefore admits a limit in distribution Q_j which may be degenerate. Because $Q_j(T_{\nu_n})$ and A_n are independent, the limit in distribution of $Q_j(T_{\nu_n})1_{A_n}$ exists and satisfies

$$\lim_{x\to\infty}\lim_n P(Q_j(T_{\nu_n}) > x,\ A_n) = P(Q_j = \infty)P(A_1).$$

By (13),

$$\lim_{x\to\infty}\liminf_n P(Q_j(T_{\tau^1_{j,n}}) > x,\ A_n) \geq P(Q_j = \infty)P(A_1).$$

Let l_j be the first l such that visit $\tau^l_{j,n}$ occur according to a policy of unlimited type with positive probability ($l_j < \infty$ iff $j \leq b$), and thus the visits $\tau^l_{j,n}$ with $l < l_j$ occur according to policies of limited type with probability one. Let $\mathcal{L}=\{\beta,\ f_\beta$ is of limited type$\}$. Then, if $l_j > 1$,

$$\begin{aligned}
&P(Q_j(T_{\tau^1_{j,n}}) - F_{\tau^1_{j,n}} > x,\ A_n)\\
&= \sum_{\beta\in\mathcal{L}}\sum_{y=1}^{\infty} P\left(F_{\tau^1_{j,n}} < y \mid Q_j(T_{\tau^1_{j,n}}) = x+y, A_n, Y_{\tau^1_{j,n}} = \beta\right)\\
&\qquad\qquad P(Q_j(T_{\tau^1_{j,n}}) = x+y, A_n, Y_{\tau^1_{j,n}} = \beta)\\
&\geq \sum_{\beta\in\mathcal{L}}\sum_{y=z}^{\infty} P(F^{\beta*} < y)P(Q_j(T_{\tau^1_{j,n}}) = x+y, A_n, Y_{\tau^1_{j,n}} = \beta)\\
&\geq \inf_{\beta\in\mathcal{L}} P(F^{\beta*} < z)P(Q_j(T_{\tau^1_{j,n}}) \geq x+z,\ A_n)
\end{aligned}$$

for any $z \geq 1$. It follows that

$$\lim_x \liminf_n P(Q_j(T_{\tau^1_{n,j}}) - F_{\tau^1_{j,n}} > x,\ A_n) \geq \inf_{\beta\in\mathcal{L}} P(F^{\beta*} < z)P(Q_j = \infty)P(A_1)$$

and, because $\lim_{z\to\infty} P(F^{\beta*} < z) = 1$ for each $\beta \in \mathcal{L}$, that

$$\lim_x \liminf_n P(Q_j(T_{\tau^1_{n,j}}) - F_{\tau^1_{j,n}} > x,\ A_n) \geq P(Q_j = \infty)P(A_1).$$

By (14), the previous inequality holds for $\lim_x \liminf_n P(Q_j(T_{\tau^2_{n,j}}) > x,\ A_n)$ and by iteration, we have that for each $l \leq l_j$

$$\lim_x \liminf_n P(Q_j(T_{\tau^l_{n,j}}) > x,\ A_n) \geq P(Q_j = \infty)P(A_1). \tag{15}$$

On the other hand, for any (j,α) and any m,

$$E\left(F_{\tau_{\tau^m_{(j,\alpha),n}}}1_{A_n}\right) \;\geq\; E(f_\alpha(x))P\left(Q_j(T_{\tau^m_{(j,\alpha),n}}) \geq x,\ A_n\right)$$

and thus

$$\liminf_n E\left(F_{\tau^m_{(j,\alpha),n}}1_{A_n}\right) \;\geq\; \liminf_x \left(E(f_\alpha(x))\liminf_n P\left(Q_j(T_{\tau^m_{(j,\alpha),n}}) \geq x,\ A_n\right)\right). \quad (16)$$

Fix $j \leq b$. We have to prove that Q_j is proper. Suppose the opposite, i.e. $P(Q_j = \infty) > 0$. Without loss of generality, suppose that policy f_1 is of unlimited type and that the event

$$A_{(j,1),n} = \{\tau^{l_j}_{j,n} = \tau^1_{(j,1),n} < \nu_{n+1}\}$$

has positive probability. For the events $A_{(j,1),n}$ and for $l = l_j$, (15) provides

$$\lim_x \liminf_n P\left(Q_j(T_{\tau^1_{(j,1),n}}) > x,\ A_{(j,1),n}\right) \;>\; 0$$

and (16) implies that

$$\liminf_n E\left(F_{\tau^1_{(j,1),n}}1_{A_{(j,1),n}}\right) \;=\; \infty$$

because $\lim_x E(f_1(x)) = E(F^{1*}) = \infty$. But $G_{j,n} \geq F_{\tau^1_{(j,1),n}}1_{A_{(j,1),n}}$ and thus $\bar{G}_j = \lim_n E(G_{j,n}) = \infty$. This is not possible because $\bar{G}_j < \infty$ by Lemma 2. Thus, Q_j is proper for $j \leq b$.

Fix now $j > b$. By Lemma 2, $\bar{G}_j < \bar{G}^*_j$. Thus there exists γ such that

$$\liminf_n E\left(\sum_{k=\nu_n}^{\nu_{n+1}-1} F_k 1_{\{(X_k,Y_k)=(j,\gamma)\}}\right) < \frac{p_{(j,\gamma)}}{p_{(1,1)}}E(F^{\gamma*})$$

with $E(F^{\gamma*}) < \infty$. Suppose Q_j concentrated at infinity, i.e. $P(Q_j = \infty) = 1$. Consider the events $A_{(j,\gamma),l,m,n} = \{\tau^l_{j,n} = \tau^m_{(j,\gamma),n} < \nu_{n+1}\}$, which are such that

$$E\left(\sum_{k=\nu_n}^{\nu_{n+1}-1} F_k 1_{\{(X_k,Y_k)=(j,\gamma)\}}\right) \;=\; \sum_{l,m} E(F_{\tau^m_{(j,\gamma),n}}1_{A_{(j,\gamma),l,m,n}}).$$

By (15), which is now valid for any l because here $l_j = \infty$,

$$\lim_x \liminf_n P(Q_j(T_{\tau^l_{j,n}}) > x,\ A_{(j,\gamma),l,m,n}) \geq P(A_{(j,\gamma),l,m,1}).$$

By (16),

$$\liminf_n E\left(F_{\tau^m_{(j,\gamma),n}}1_{A_{(j,\gamma),l,m,n}}\right) \;\geq\; E(F^{\gamma*})P(A_{(j,\gamma),l,m,1}).$$

It implies that

$$\begin{aligned}\liminf_n E\left(\sum_{k=\nu_n}^{\nu_{n+1}-1} F_k 1_{\{(X_k,Y_k)=(j,\gamma)\}}\right) &\geq \sum_{l,m} \liminf_n E\left(F_{T^m_{(j,\gamma),n}} 1_{A_{(j,\gamma),l,m,n}}\right). \\ &\geq E(F^{\gamma *}) \sum_{l,m} P(A_{(j,\gamma),l,m,1}) \\ &= E(F^{\gamma *}) \frac{p_{(j,\gamma)}}{p_{(1,1)}}\end{aligned}$$

This is a contradiction. Thus Q_j cannot be concentrated at infinity. □

The next theorem is our main result.

Theorem 2 *Condition $\mathcal{C}$ is necessary and sufficient for the stability of the polling system.*

Proof: The strategy of proof is as in [4]. First we prove the sufficiency. For each e, the multidimensional Markov chain $(M^e_{\nu_n})_n$ is ergodic iff m^e exists such that $\lim_{n\to\infty} P(M^e_{\nu_n} \leq m^e) > 0$. On the other hand,

$$\lim_{n\to\infty} P(M^e_{\nu_n} \leq m^e) \geq 1 - \sum_{k=1}^{e} \lim_{n\to\infty} P(Q^e_k(T^e_{\nu_n}) \geq m^e_k)$$

and the sum on the right hand side can be made less than one if the limiting distribution of $Q^e_k(T^e_{\nu^e_n})$ is not concentrated at infinity for one k and is proper for the others.

We proceed by induction on the subsystems. First, consider $\mathcal{S}^b$ with $\hat{\rho}^b < 1$, $(X_1, Y_1) = (1,1)$ and $M^b_1 = (0,\ldots,0)$. By Lemma 3, each component of $M^b_{\nu_n}$ has a proper limit in distribution. Thus $M^b_{\nu_n}$ is ergodic and converges in distribution to its stationary distribution independently of the initial state. By Lemma 2, for each $j \leq b$, $\bar{G}^b_j < \infty$. By equation (8), the mean cycle times converge monotonically to a finite limit. The stationary cycle time is thus integrable. By standard arguments on Markov chains, the system admits integrable regeneration times, for example the times of return of $M^b_{\nu_n}$ to $(0,\ldots,0)$. Hence $\mathcal{C}^b$ is sufficient for the stability of $\mathcal{S}^b$.

Suppose now $\mathcal{C}^{e-1}$ sufficient for the stability of $\mathcal{S}^{e-1}$ and consider $\mathcal{S}^e$ with $(X_1, Y_1) = (1,1)$ and $M^e_1 = (0,\ldots,0)$ $(e > b)$. Suppose that $\mathcal{C}^e$ holds. It implies that $\mathcal{C}^{e-1}$ holds and, therefore, that $(M^{e-1}_{\nu_n})_n$ converges to a proper limit in distribution independently of the initial state. Because for each n $M^{e|e-1}_{\nu_n} \leq_d M^{e-1}_{\nu_n}$ (Lemma 1), $(M^{e|e-1}_{\nu_n})_n$ admits a proper limit in distribution too. By Lemma 3, the limit in distribution of $Q^e_e(T^e_{\nu_n})$, the last component of $M^e_{\nu_n}$, is not concentrated at infinity. Thus $M^e_{\nu_n}$ is ergodic and, as above, the stationary cycle time is integrable. Condition $\mathcal{C}^e$ is sufficient for the stability of $\mathcal{S}^e$. By induction, the proof of sufficiency is complete.

For the necessity, suppose $\mathcal{S}$ stable with (X_n, Y_n, M_n) stationary. Here we consider cycles with respect to a state at which queue c is served, say (c,δ) with $p_{(c,\delta)} > 0$, and adapt the previous calculations to (c,δ)-cycles. The cycle times are stationary and are integrable. Thus the mean number of customers served in queue j in (c,δ)-cycle n does not depend on n and is finite for each j, i.e. $E(G_{j,(c,\delta),n}) = \bar{G}_{j,(c,\delta)} < \infty$. From equation (5) transposed to (c,δ)-cycles and for each j,

$$-G_{j,(c,\delta),1} \leq Q_j(T_{\nu_{(c,\delta),2}}) - Q_j(T_{\nu_{(c,\delta),1}}) \leq N_j(T_{\nu_{(c,\delta),1}}, T_{\nu_{(c,\delta),2}}]$$

where both bounds are integrable. Thus $(Q_j(T_{\nu_{(c,\delta),n}}))_n$ is stationary with integrable increments and, by Lemma 7 of [4], the increments have zero expectation:

$$E(Q_j(T_{\nu_{(c,\delta),n+1}}) - Q_j(T_{\nu_{(c,\delta),n}})) = 0.$$

This transforms (10) into the system of equations

$$\bar{G}_{j,(c,\delta)} = \lambda_j \left(\sum_{k=1}^{c} \sigma_k \bar{G}_{k,(c,\delta)} + \frac{\bar{S}}{p_{(c,\delta)}} \right) \qquad (1 \leq j \leq c) \tag{17}$$

and (11) becomes

$$(1-\hat{\rho}^b) \sum_{j=1}^{b} \sigma_j \bar{G}_{j,(c,\delta)} = \hat{\rho}^b \left(\sum_{j=b+1}^{c} \sigma_j \bar{G}_{j,(c,\delta)} + \frac{\bar{S}}{p_{(c,\delta)}} \right).$$

Thus $\hat{\rho}^b < 1$ because the right hand side is positive. The proof is complete when $b = c$. When $b < c$, we also get equalities in (12):

$$(1-\hat{\rho}^j)\bar{G}_{j,(c,\delta)} = \lambda_j \left(\sum_{k=j+1}^{c} \sigma_k \bar{G}_{k,(c,\delta)} + \frac{\bar{S}}{p_{(c,\delta)}} \right) \qquad (b < j \leq c).$$

The last equation provides

$$\bar{G}_{c,(c,\delta)} = (1-\hat{\rho}^c)^{-1} \lambda_c \frac{\bar{S}}{p_{(c,\delta)}}.$$

On the other hand,

$$G_{c,(c,\delta),1} = F_{\nu_{(c,\delta),1}} + \sum_{\beta \neq \delta} \sum_{l=\nu_{(c,\delta),1}}^{\nu_{(c,\delta),2}-1} 1_{\{(X_1,Y_1)=(c,\beta)\}} F_l.$$

Because $P(Q_c(T_{\nu_{(c,\delta),1}}) = 0) > 0$ by the ergodicity of $(M_{\nu_{(c,\delta),n}})_n$ and because policy f_δ is of limited type (here $b < c$), $E(F_{\nu_{(c,\delta),1}}) < E(F^{\delta *})$ which implies that

$$\bar{G}_{c,(c,\delta)}. < E(F^{\delta *}) + \sum_{\beta \neq \delta} \frac{p_{(c,\beta)}}{p_{(c,\delta)}} E(F^{\beta *}).$$

Thus,

$$(1-\hat{\rho}^c)^{-1} \lambda_c \bar{S} < \sum_{\beta=1}^{\delta} p_{(c,\beta)} E(F^{\beta *})$$

which is equivalent to condition $\mathcal{C}$. $\square$

Remarks. 1. The mean stationary number of customers served in queue j per visit is easily obtained from the solution of (17) and equals $\lambda_j \bar{S}/(1-\hat{\rho})$.
2. As corollary, for each $b \leq e \leq c$, condition $\mathcal{C}^e$ is necessary and sufficient for the stability of polling (sub)system $\mathcal{S}^e$.

Examples 1. When $(X_., Y_.)$ is deterministic periodic, the trajectory is a repeating sequence of $(t(i), i)_{i=1,\ldots,\delta}$ where $t(i) = j \in \{1, \ldots, c\}$ means that queue j is served according to policy f_i. Then, $p_{(j,i)}$ equals $1/\delta$ if $j = t(i)$ and zero otherwise, and the stability condition is

$$\hat{\rho} + \max_j \left(\frac{\lambda_j}{\sum_{l=1}^{a_j} E(F^{j_l *})} \right) S < 1$$

where $S = \sum_{i=1}^{\delta} S_{(t(i),i),(t(i+1),i+1)}$ is the mean switchover time in a cycle and where j_l indexes the a_j visits to queue j in a cycle, i.e. $t^{-1}(j) = \{j_1, \ldots, j_{a_j}\}$. This is the condition obtained in [4].
2. When X_n is a Markov chain and Y_n is a deterministic function of X_n, we have the Markovian routing with a fixed service policy for each queue, say f_j for queue j. Then $p_{(j,\alpha)} > 0$ iff $f_\alpha = f_j$ and the stability condition is

$$\hat{\rho}^c + \max_{1 \leq j \leq c} \left(\frac{\lambda_j}{p_{(j,j)} E(F^{j*})} \right) \bar{S} < 1.$$

When policy f_j is a limited gated policy, $f_j(x) = \min(x, L_j)$ with $E(L_j) < \infty$, the condition has been derived in [2].

The analysis allows to compare the queues in terms of stability in case of heavy traffic. Suppose that the system is stable and that the arrival rates to the queues increase in the same proportion. Then, the order in which the queues become unstable is the decreasing order $\lambda_j / \sum_{\alpha=1}^{\delta} p_{(j,\alpha)} E(F^{\alpha *})$, simultaneously in case of equality. The analysis also allows to determine the local stability conditions of the queues, under which they are stable even when the whole system is not. Indeed, the queues $1, \ldots, b$, are simultaneously stable or simultaneously unstable according to $\hat{\rho}^b < 1$ or not, respectively, and queue j $(j > b)$ is stable if and only if

$$\hat{\rho}^j + \frac{\lambda_j}{\sum_{\beta=1}^{\delta} p_{(j,\beta)} E(F^{\beta *})} \bar{S}^j < 1.$$

When $\hat{\rho}^b < 1$ (if not, each queue is unstable), define κ as the greatest e such that $\mathcal{C}^e$ holds. When $\kappa < c$, i.e. the system is unstable, queues $1, \ldots, \kappa$ are stable while the other queues are not.

Finally, when each queue is eventually served according to a policy of unlimited type, i.e. when $b = c$, the condition $\hat{\rho}^b < 1$ is also necessary and sufficient for the stability when set-up times of the server are added: at each visit to a nonempty queue the server needs an amount of time before starting to serve. Moreover the set-up times and the switchover times can be state-dependent in a stochastic monotonic way, for example monotonic in the queue length at the polling time and in the number of served customers, respectively, as long as they are stochastically dominated by an integrable random variable.

These results on local stability and on the model with set-up times are treated with more details for the periodic polling model in [4] .

References

[1] E. Altman, P. Konstantopoulos and Z. Liu, Stability, Monotonicity and Invariant Quantities in General Polling Systems, Queueing Systems 11 (1992) 35-58.

[2] A.A. Borovkov and R. Schassberger, Ergodicity of a polling network, Techn. Report (1993).

[3] M. Eisenberg, Queues with Periodic Service and Changeover Time, Oper. Res. 20 (1972) 440-451.

[4] C. Fricker and M.R. Jaïbi, Monotonicity and stability of periodic polling Models, Queueing Systems 15 (1994) 211-238.

[5] L. Georgiadis and W. Szpankowski, Stability of Token Passing Rings, Queueing systems 11 (1992) 7-34.

[6] P.J. Kuehn, Multiqueue Systems with Nonexhaustive Cyclic Service, The Bell System Technical Journal 58 (1979) 671-698.

[7] L. Massoulié, Stability of non Markovian Polling Systems, INRIA Report 2148 (1993).

[8] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley & Sons (1983).

[9] H. Takagi, Queueing Analysis of Polling Systems, in: *Stochastic Analysis of Computer and Communication Systems,* H. Takagi, ed. (North Holland, Amsterdam, 1990) 267-318.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine - Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes - IRISA, Campus universitaire de Beaulieu 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 Grenoble Cedex 1 (France)
Unité de recherche INRIA Sophia Antipolis - 2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249 - 6399

RR_2278